"VIDEO ENCODING METHOD"

FIELD OF THE INVENTION

The present invention generally relates to the field of object-based video coding, and more particularly to a method for encoding a sequence of video data according to a process in which several types of data are identified, said data consisting of so-called Video Object Planes (VOPs) that are either intra coded VOPs (I-VOPs), coded using information only from themselves, or predictive coded VOPs (P-VOPs), coded using a motion compensated prediction from a past reference VOP, or bidirectionally predicted VOPs (B-VOPs), coded using a motion-compensated prediction from past and future reference VOPs.

BACKGROUND OF THE INVENTION

The MPEG-4 visual standard provides technologies to view, access and manipulate objects (rather than pixels, in the case of the previous MPEG standards) in a large range of bit rates, in a lot of application areas such as for instance : digital television, streaming video, mobile multimedia, games, etc. Said standard operates on video objects (VOs) defined by temporal and spatial information in the form of shape, motion and texture information, coded separately in the bitstream (these VOs are the entities that the user can access and manipulate).

The MPEG-4 approach relies on a content-based visual data representation of the successive scenes of a sequence, each scene being a composition of VOs with its intrinsic properties : shape, motion, texture. In addition to the concept of VO, the MPEG-4 standard introduces other ones like the Video Object Layer (each VO can be encoded either in a scalable or non-scalable form, depending on the application, represented by the video object layer, or VOL) and the Video Object Planes (VOPs) (= instances of VOs in time). It is assumed that each frame of an input video sequence is segmented into a number of arbitrarily shaped image regions (the VOs), and that the shape, motion and texture information of the VOPs belonging to the same VO is encoded and transmitted into separate VOLs corresponding to specific temporal or spatial resolutions (which allows later to separately decode each VOP and leads to the required flexible manipulation of the video sequence).

The three types of frames processed by such a coding structure are the following : the I-VOPs, the P-VOPs and the B-VOPs. An I-VOP is an intra coded VOP : the coding operation uses information only from itself (it is the VOP that costs the greatest number of bits). A P-VOP is a predictive coded VOP, and the coding operation then uses a motion

compensated prediction from a past reference VOP which can be either an I-VOP or another P-VOP (contrary to an I-VOP, only the difference between the current motion-compensated P-VOP and its reference is coded : thus, a P-VOP usually costs fewer bits than an I-VOP). A B-VOP is a VOP that is coded using a motion-compensated prediction from past and future

5   reference VOPs (I or P-VOPs), based on so-called forward and backward motion estimations respectively. A B-VOP cannot be a reference VOP and, like the P-VOP, only the difference between the current motion compensated B-VOP and its reference VOP is coded.

Unfortunately, using said B-VOP prediction (also called interpolated or bi-directional mode) is not always a gain in term of compression. If the compression can sometimes

10 be improved by a factor of about 20 %, it can also in other cases be decreased by a drastic factor.

SUMMARY OF THE INVENTION

It is then an object of the invention to propose an encoding method using this B-VOP prediction only when it is efficient.

To this end, the invention relates to an encoding method such as defined in

15 the introductory part of the description, said encoding method including a coding step of each VOP and, before said coding step, a motion estimation step performed between the current VOP and the previous one, said motion estimation step itself comprising a decision process concerning the type of VOP to be coded and based on the sub-steps of :

- carrying out a motion estimation between a VOP number N (VOP N) and

20 the previous one (VOP N-1) ;

- on the basis of said motion estimation, computing a so-called coherence factor, provided for quantifying the sequence motion ;

- on the basis of a comparison of said coherence factor with a predetermined threshold, taking a final decision on the type of the current VOP, said current VOP being a B-

25 VOP or not according to the value of said coherence factor with respect to said threshold.

BRIEF DESCRIPTION OF DRAWINGS

The present invention will now be described, by way of example, with reference to the accompanying drawings in which Fig.1 illustrates the main steps of the encoding method according to the invention.

30 **DETAILED DESCRIPTION OF THE INVENTION**

An MPEG-4 encoder comprises several functional blocks, among which one or several memories, for outputting the VOPs in the transmitting order required by the standard.

For example, if the input sequence is I B B P B B P..., the output or transmission order will be I P B B P B B.... (to enable backward prediction, the encoder must re-order the pictures from natural order to said transmission order so that the B-VOPs are transmitted after the past and future VOPs which are their references). Said encoder also comprises a motion estimator, for

5   receiving the current VOP and the previous one (or reference VOP) and taking the decision of which kind of prediction will be implemented for the current VOP : no prediction for an I-VOP, forward prediction for a P-VOP, bi-directional prediction for a B-VOP.

As illustrated in Fig.1 showing the steps of the decision about the prediction within the motion estimator, the decision is, according to the invention, based on the

10 implementation of the following steps. First, the current VOP (number : N) is captured (step CAP-VN). Then a motion estimation is carried out (step M-EST) between the VOP N and the previous VOP (number : N-1), and a factor named "coherence factor" is computed (step CF-CPT), in order to quantify the sequence motion, and compared (step CF-TEST) to a predetermined threshold. According to the result of the comparison (Y = lower than, or N = not

15 lower than), the VOP N is allowed (N-AL), or not (N-NAL), to be a B-VOP. The final decision concerning the prediction mode is then taken (step V-DEC), and the coding step of the current VOP (= I-VOP, or P-VOP, or B-VOP) can take place (step COD).

Several expressions may be proposed for the coherence factor used in the comparison test, without limiting the scope of the invention. However, a preferred coherence

20 factor may be for instance expressed as the ratio of the sum of absolute differences (SAD) between motion vectors of a macroblock (estimated in 16 x 16 pixels mode or 8 x 8 pixels mode) and its predecessor in the same VOP with the similar sum for the previous VOP (it is here recalled that for a macroblock of size k x k, the expression of the SAD is :

$$SAD = \sum_{i=0}^{kxk} |A(i) - B(i)|$$

25 where B(i) and A(i) respectively designate the current macroblock considered and the macroblock in the reference VOP which matches the most in a search zone defined in said reference VOP).